Instantaneous Higher Order Phase Derivatives

Douglas J. Nelson

National Security Agency, Fort George G. Meade, Maryland 20755
E-mail: djnelso@afterlife.ncsc.mil


We present methods, based on the short time Fourier transform, which may be used to analyze the structure of multicomponent FM modulated signals instantaneously in time and frequency. The methods build on previously presented cross-spectral methods. In this paper, we introduce the concept of higher order short time Fourier transform phase derivatives, which may be used to estimate signal trajectories instantaneously in both time and frequency and to determine convergence of the remapped time–frequency surface. The methods are applied to synthesized data and speech signals.

Key Words: cross-spectrum; phase-spectrum; spectral estimation; short time Fourier transform; time–frequency representation.

0. INTRODUCTION

Since the short time Fourier transform (STFT) and the spectrogram were introduced by Gabor to study the local time–frequency (TF) properties of signals [5], there have been many methods developed to estimate the instantaneous characteristics of signals. While other methods, such as the Wigner distribution, wavelet transform, and variations of these methods, have become quite popular, Fourier transform-based methods such as the spectrogram still remain the standard. Among the advantages of the Fourier methods are the efficiency of computation and the fact that the Fourier transform is a linear process, which provides an unbiased spectral estimate. The main disadvantage of Fourier representations is that they generally produce poor resolution in time and frequency. While nonlinear techniques, such as the Wigner distribution, in principle, produce better TF resolution than the Fourier representations, they also have undesirable properties, such as unwanted intermodulation products or cross-terms [3]. Much of the current TF literature is dedicated to modifying the Wigner distribution to mitigate the cross-terms and make the modified distribution appear more like an energy density function (cf. [1–3, 8]). In this work, we present results based on previously reported cross-spectral methods which
have been used to produce high resolution TF representations from the phase of the STFT. The methods are inspired by instantaneous frequency (IF), which is the derivative of the phase of the analytic signal [17].

If the signal is known to consist of a single FM component, the IF provides a good TF representation of the signal. A method based on the IF was used by Kay [6] to accurately estimate an isolated sine wave in noise. The problem is that the simple IF method will not work on multicomponent signals. In a series of papers by Nelson and others, the cross-power spectrum has been used to accurately estimate multiple stationary sine waves in noise (cf. [10, 11, 18]). The cross-power spectrum (CPS) is actually a TF method for simultaneously estimating the frequencies of multiple sine waves in deep noise. It is based on the derivative, with respect to time, of the STFT phase. In Monte Carlo runs, it was verified that a variation of the CPS algorithm provides unbiased frequency estimates of sine waves in noise, with performance attaining the Cramer–Rao bound [18].

These works, and works such as Friedman’s [4], represent early attempts to estimate multicomponent signals from the phase differentiated STFT surface. In applying these methods to nonstationary signals such as speech, the approach has generally been to assume that the signal may be approximated as the sum of piecewise stationary sine waves. Recently cross-spectral methods based on the combined use of the channelized instantaneous frequency (CIF) and the local (LGD) functions, which we define in the next section, have been proposed as a method of remapping the STFT surface and estimating the instantaneous properties of nonstationary multicomponent FM signals jointly in time and frequency [12, 14, 16]. It should be noted that the use of these two derivatives was actually proposed by Kodera in 1976 [7]. In the case of linear FM signals, these methods have been proven analytically and verified experimentally to correctly remap the entire STFT surface to a single line, which agrees with the IF representation of the signal [16]. The cross-spectral methods may be used to accurately estimate the instantaneous TF properties of multicomponent signals, whose components may be approximated as piecewise linear FM. Unlike the methods based only on the CIF representation, which approximate nonstationary signals locally as sine waves, cross-spectral remapping is based jointly on instantaneous frequency and group delay and does not make any use of a stationarity assumption.

We briefly describe previously reported cross-spectral methods and extend the methods by introducing higher order phase derivative surfaces. We provide methods for accurately estimating them, and we present interpretations of two of these surfaces. The methods are applied to synthesized data and speech signals, demonstrating their utility in signal processing.

This paper is structured as follows. In Section 1, we define the STFT, phase derivatives, their computation as cross-spectral products, and the use of spectral phase in surface remapping. In Section 2 we develop a convergence condition for surface remapping based on the mixed partial derivatives of the STFT phase. In Section 3 we introduce the concept of instantaneous slope as a TF surface and
relate the local properties of this surface to the individual FM components which comprise the signal. In Section 4, we describe experimental results.

1. THE STFT AND SURFACE DERIVATIVES

The STFT was introduced by Gabor in 1946 as a method of studying the behavior of signals jointly in time and frequency [5]. In the STFT, the Fourier transforms of the product of the signal \( f(t) \) and a sequence of time translations of a (short) analysis window \( h(-t) \) are computed. The STFT may therefore be represented as

\[
F(\omega, T) = \int_{-\infty}^{\infty} f(t + T) h(-t) e^{-i\omega t} dt. \tag{1}
\]

While \( F(\omega, T) \) is dependent on the windowing function \( h(-t) \), we drop the \( h \) from the notation for simplicity. It can easily be seen that \( F(\omega_0, T) \) is the convolution of \( f(t) \) and \( h(t) e^{i\omega_0 t} \). For fixed \( \omega_0 \), \( F(\omega_0, T) \) is therefore a (bandpass) filtered version of the original signal [9]. If \( h(t) e^{i\omega_0 t} \) is analytic, \( F(\omega_0, T) \) is easily seen to be analytic, even if \( f(t) \) is real [16].

1.1. Surface Remapping

We briefly describe the cross-spectral remapping method, which is based on the signal IF and time and frequency error estimates recovered from the CIF and LGD surfaces. These functions may be represented as

\[
\text{IF}_f(t) = \frac{d}{dt} \arg\{f(t)\} \quad (2)
\]

\[
\text{CIF}_f(\omega, T) = \frac{d}{dT} \arg\{F(\omega, T)\} \quad (3)
\]

\[
\text{LGD}_f(\omega, T) = \frac{d}{d\omega} \arg\{F(\omega, T)\}, \quad (4)
\]

where we have assumed the analytic representation of the signal in the IF representation.

Assume a signal which may be written as the sum of AM and FM modulated components

\[
\sum_n f_n(t), \quad (5)
\]

where \( f_n(t) = A_n(t)e^{i\Omega_n(t)}, \omega_n(t) = \Omega'_n(t), A_n(t) \) is continuous and positive, and \( \Omega_n(t) \) has continuous derivatives of whatever order we require. We would like a method which isolates and simultaneously estimates each of the signal components \( f_n(t) \). We assume that the IF representation of each signal component accurately represents the TF behavior of that component. The goal therefore is to effectively derive the IF representations of each of the signal components from the STFT. It has been demonstrated that signal components, which agree with our intuitive notion of single component, can be accurately estimated directly from the STFT, as long as the surface satisfies a reasonable local separability condition [13, 16].
Separability is simply the condition that, for any point \((\omega_0, T_0)\), at most one signal component can contribute a significant amount of energy to the STFT surface in a neighborhood of that point. Specifically, we define \(F(\omega, T)\) to be separable at \((\omega_0, T_0)\), if, for some \(n\),

\[
|F_n(\omega, T)|^2 \gg \sum_{m \neq n} |F_m(\omega, T)|^2,
\]

where \(F_n(\omega, T)\) is the STFT of the signal component \(f_n(t)\), and we assume that \(|\omega - \omega_0| < \epsilon\) and \(|T - T_0| < \delta\), for some small \(\epsilon\) and \(\delta\).

Separability guarantees that the STFT surface and all of its derivatives locally represent the STFT and derivatives of the strongest signal components. If \(F_n\) is the dominant signal component at \((\omega_0, T_0)\), then \(\text{CIF}_f(\omega_0, T_0)\) provides a re-estimation of the frequency of that component observed at \((\omega_0, T_0)\), and \(\text{LGD}_f(\omega_0, T_0)\) represents an estimated timing error of the observation. The relationship

\[
(\omega, T) \rightarrow (\text{CIF}_f(\omega, T), T + \text{LGD}_f(\omega, T))
\]

(7)

provides a remapping of the TF plane, under which separable points on the STFT surface are remapped to points along curves represented by the IF representations of the individual FM components of the signal [16]. That is,

\[
\text{CIF}_f(\omega_0, T_0) \approx \text{IF}_{f_n}(T_0 + \text{LGD}_f(\omega_0, T_0)),
\]

(8)

where \(f_n\) is the dominant signal component at the separable point \((\omega_0, T_0)\). In effect, remapping the surface concentrates the surface components along curves which agree with the IF representation of the signal

\[
\omega_n(T) = \text{IF}_{f_n}(T).
\]

(9)

Equation (8) provides a simple relationship among the first-order derivatives. We would like to establish conditions which measure convergence and accuracy of the remapping. In addition, we would like to estimate higher order phase derivatives of the FM signal components directly from the STFT of the composite signal.

1.2. Phase Derivatives as Cross-Spectral Products

We define the following notation for derivatives of the signal and STFT phase

\[
\phi_f^{[n]}(t) = \frac{\partial^n}{\partial t^n} \arg\{f(t)\}
\]

(10)

\[
\Phi_f^{[\Psi]}(\omega, T) = \frac{\partial^n}{\partial \psi_n \cdots \partial \psi_1} \arg\{F(\omega, T)\},
\]

(11)

where \(\Psi = (\psi_1, \ldots, \psi_n)\) and the \(\psi_k\) assume the values \(\omega\) or \(T\).
The first-order derivatives have convenient interpretations as the IF, CIF, and LGD functions

\[
\text{IF}_f(t) = \phi^1_f(t) = \frac{d}{dt} \arg\{f(t)\} \tag{12}
\]

\[
\text{CIF}_f(\omega, T) = \Phi^1_f(\omega, T) = \frac{d}{dT} \arg\{F(\omega, T)\} \tag{13}
\]

\[
\text{LGD}_f(\omega, T) = -\Phi^1_f(\omega, T) = -\frac{d}{d\omega} \arg\{F(\omega, T)\}, \tag{14}
\]

where we assume that \(\arg\{F(\omega, T)\}\) is continuously differentiable. To estimate the CIF and LGD and the higher order derivative surfaces, we define the cross-spectral surfaces

\[
F^{[\delta_T]}(\omega, T) = F\left(\omega, T + \frac{\delta_T}{2}\right) F^*\left(\omega, T - \frac{\delta_T}{2}\right) \tag{15}
\]

\[
F^{[\delta_\omega]}(\omega, T) = F\left(\omega + \frac{\delta_\omega}{2}, T\right) F^*\left(\omega - \frac{\delta_\omega}{2}, T\right) \tag{16}
\]

\[
F^{[\Delta, \delta_T]}(\omega, T) = F^{[\Delta]}\left(\omega, T + \frac{\delta_T}{2}\right) F^{*[\Delta]}\left(\omega, T - \frac{\delta_T}{2}\right) \tag{17}
\]

\[
F^{[\Delta, \delta_\omega]}(\omega, T) = F^{[\Delta]}\left(\omega + \frac{\delta_\omega}{2}, T\right) F^{*[\Delta]}\left(\omega - \frac{\delta_\omega}{2}, T\right), \tag{18}
\]

where \(\Delta = (\delta_{\phi_1}, \ldots, \delta_{\phi_n})\) and the \(\delta_{\phi_k}\) represent small intervals in \(T\) or \(\omega\).

The spectrogram (squared magnitude of the STFT surface), LGD, and CIF surfaces may then be estimated as

\[
|F(\omega, T)|^2 \approx |F^{[\delta_T]}(\omega, T)| \approx |F^{[\delta_\omega]}(\omega, T)| \tag{19}
\]

\[
\text{CIF}_f(\omega, T) \approx \frac{1}{\epsilon} \arg\{F^{[\delta_T]}(\omega, T)\} \tag{20}
\]

\[
\text{LGD}_f(\omega, T) \approx -\frac{1}{\epsilon} \arg\{F^{[\delta_\omega]}(\omega, T)\}. \tag{21}
\]

The higher order phase derivatives may be estimated as

\[
\Phi^1_f(\omega, T) \approx \arg\left\{ \prod_{k=1}^n \frac{1}{\delta_{\phi_k}} F^{[\Delta]}(\omega, T) \right\}. \tag{22}
\]

Equation (22) provides a simple method of estimating the higher order phase derivatives directly from a sampling lattice of a discrete STFT surface.

2. MIXED PARTIAL CONDITION FOR CONVERGENCE

The remapping Eq. (7) provides a redistribution of the entire TF surface, but the remapping is point-wise, and it is only valid at separable points, where one of the signal components has significant energy. For these points, remapping concentrates the surface energy along the curves of Eq. (8) corresponding to the IF representations of the individual signal components. Under remapping, regions of the STFT surface where the signal has little energy are randomly...
mapped, resulting in considerable speckled, low energy noise. Many signals, such as speech, have relatively slowly varying spectral components. For speech, both the glottal excitation frequency and the resonant frequencies (formants) normally have a rate of change which is less than 500 Hz/s. For such signals, there is a simple convergence test based on the mixed partial derivatives of phase. This test may be quite effective in removing the noise speckle from the remapped STFT.

To derive the test, we simply note that the mixed partial phase derivative may be represented as

$$\frac{\partial^2}{\partial \omega \partial T} \arg F(\omega_0, T_0) \approx \frac{1}{\epsilon} \left( \text{CIF}_f \left( \frac{\omega_0 + \epsilon}{2}, T_0 \right) - \text{CIF}_f \left( \frac{\omega_0 - \epsilon}{2}, T_0 \right) \right)$$

(23)

for small $\epsilon$. The right-hand side of Eq. (23) is precisely the ratio of the frequency difference of the two points on the remapped surface to the frequency difference of the two points on the original STFT surface. The mixed partial phase derivative of Eq. (23) is therefore a measure of the local coalescence in frequency of the remapping. Equivalently, it is a measure of how well the remapped surface approximates a stationary signal component. To use Eq. (23) to remove noise speckle from the remapped surface, we simply set a convergence threshold and discard remapped points whose mixed phase partial derivative exceed this threshold. This convergence test was applied as part of a speech pitch (excitation frequency) estimation algorithm. On more than 400 files from the SWITCHBOARD database, it was effective in removing nearly all the noise by simply discarding all data for which the value of the mixed partial phase derivative exceeded 0.5.

We may apply the same logic to the mixed partial derivative

$$\frac{\partial^2}{\partial \omega \partial T} \arg F(\omega_0, T_0) \approx \frac{1}{\epsilon} \left( \text{LGD}_f \left( \frac{\omega_0 + \epsilon}{2}, T_0 + \text{LGD}_f (\omega, T) \right) - \text{LGD}_f \left( \frac{\omega_0 - \epsilon}{2}, T_0 - \text{LGD}_f (\omega, T) \right) \right).$$

(24)

The convergence test resulting from Eq. (24) is a measure of how localized the signal component is in time. If the mixed partial derivative equals unity, the remapped surface converges locally to an impulse in time. It should be noted that, in estimating the mixed partial derivatives as cross-spectral products, the order of differentiation may be interchanged, so it is not necessary to compute two sets of derivatives.

3. INSTANTANEOUS STFT SLOPE

For the case of a single linear FM signal component, it was proven that the entire STFT surface theoretically remaps to the single line which agrees with the IF representation of the signal [16]. That is, for a linear FM signal and any point $(\omega, T)$, the following relationship is valid

$$\text{CIF}_f (\omega, T) = \text{IF}_f (T + \text{LGD}_f (\omega, T)).$$

(25)
For nonlinear components, which may be piece-wise approximated by linear FM segments, we may expect STFT surface components near the IF representation to be remapped to the IF representation in the TF plane. That is, for FM representations, which may be linearly approximated, we should expect Eq. (25) to be satisfied. This relationship has indeed been verified on a large number of synthesized signals, as well as speech, marine mammal sounds, and other real signals [12, 14–16].

From Eq. (25), we can estimate the instantaneous frequency \( f(t) \) of signal components which are separable on the STFT surface. It is also possible to directly estimate the higher order signal phase derivatives \( \phi^{[n]}(t) \) from the remapped STFT. We will demonstrate this for the instantaneous slope, \( \phi^{[2]}(t) \), of the FM component. We select two points within a small neighborhood on the STFT surface and calculate their images under remapping. Assuming separability, if the neighborhood is sufficiently small, and the STFT surface energy at each of the two points is sufficiently large, we may assume that both points are attracted to the same FM component. We may therefore expect each of the selected points to be remapped to an estimated point along the IF representation of the signal component. That is, for two points, \((\omega_0, T_0)\) and \((\omega_1, T_1)\), such that \(|\omega_1 - \omega_0| < \delta\) and \(|T_1 - T_0| < \epsilon\), for suitably small \(\delta\) and \(\epsilon\), and

\[
|F(\omega_0, T_0)| \gg 0, \quad |F(\omega_1, T_1)| \gg 0, \tag{26}
\]

we may expect the remapped points \((\text{CIF}_{f}(\omega_k, T_k), T_k + \text{LGD}_{f}(\omega_k, T_k))\) to approximately lie on the IF representation of the signal. The slope of the line segment between the two remapped points therefore represents an estimate of the second derivative of the phase of the dominant analytic signal component, or the instantaneous slope of its IF representation, at some time

\[
t \in (T_0 + \text{LGD}_{f}(\omega_0, T_0), T_1 + \text{LGD}_{f}(\omega_1, T_1)). \tag{27}
\]

The instantaneous slope (IS) is a TF surface, which we may represent as

\[
\text{IS}_{f}(\omega, T, \alpha) = \frac{\partial_{\alpha} \text{CIF}_{f}(\omega, T)}{\partial_{\alpha}(T + \text{LGD}_{f}(\omega, T))} \tag{28}
\]

where \(\partial_{\alpha}\) represents the directional derivative in the \(\alpha\) direction. In principle the IS should be independent of \(\alpha\), but in practice, the variance of the estimate is dependent on \(\alpha\). Higher order derivatives of the IF representation of dominant signal components may be estimated in a similar manner.

4. TESTING AND EVALUATION

While it is not primarily the focus of this paper, the properties of the remapped STFT have been tested on a wide variety of signals, both synthesized and natural. For signals consisting of multiple sine waves in noise, the results are documented in the literature. In Fig. 1, we present a representative example of
the performance of the remapping of a single Fourier transform (i.e., $F(\omega, T_0)$).

Figure 1a represents the full spectrum (dB) and remapped spectrogram (dB). Figure 1b represents an expansion of the frequency band around one of the sine waves, showing a large reduction of the effective signal bandwidth of the signal. Figure 2 represents the formant structure of voiced speech computed

---

**FIG. 1.** Spectrum of two sine waves. Solid line, spectrogram (dB). Dotted line, remapped STFT (dB). Top trace, entire spectrum. Bottom trace, expansion of one spectral bulge.

**FIG. 2.** Averaged speech formant spectra (TIMIT). Each trace computed by averaging spectra computed from $1/200$ s of data. Total integration time: $1/10$ s. Top trace, averaged spectrogram (dB) and averaged re-mapped STFT (dB). Bottom trace, averaged spectrogram (dB) and averaged re-mapped STFT (dB) with mixed partial phase derivative threshold applied.
FIG. 3. Remapped speech excitation fundamental. Dots on rectangular grid represent STFT lattice points which survived the mixed partial test. Only these points were used in computing the remapped excitation spectrum.

as time-averaged power spectra (dB). Superimposed are the averages of the cross-spectra modified by the mixed partial phase derivative test (i.e., discarding all data for which the value of the mixed partial phase derivative magnitude exceeded 0.5), demonstrating the effectiveness of the test. Figure 3 is a portion of the STFT surface showing the points which pass the mixed partial test (i.e., points with mixed partial magnitude > 0.5 discarded) and their remapped images. Figure 4 represents the speech excitation frequency estimated as the remapped STFT points which pass the mixed partial test. These remapped points are superimposed on the spectrogram, demonstrating that the remapping has converged to the correct function. In testing the mixed partial derivative stationarity criteria, this test was used as a noise speckle filter in the pitch excitation estimation algorithm. In processing 412 speech files, nearly all of the speckle was removed, while removing almost none of the remapped fundamental, as validated by hand checking.

Figures 5–10 demonstrate the estimation of the IF and IS on a linear FM signal (chirp). The signal consisting of 2048 real samples was generated using

FIG. 4. Remapped speech excitation fundamental (white dots) superimposed on spectrogram (SWITCHBOARD database).
FIG. 5. Spectrogram of a chirp with remapped STFT (white *s) superimposed.

FIG. 6. An expansion of a portion of Fig. 5. Block grayscale structure represents TF quantization of the spectrogram.

FIG. 7. Angular frequency of a linear FM (chirp) estimated by differentiating the phase of the reconstructed analytic signal with remapped STFT superimposed.
FIG. 8. An expansion of a portion of Fig. 7. Remapped STFT points are represented by ‘s with white centers.

FIG. 9. Instantaneous slope of a linear FM (radians/sample) estimated as the second derivative of the phase of the reconstructed analytic signal with cross-spectral IS superimposed.

FIG. 10. An expansion of Fig. 9. Black ‘s with white centers represent instantaneous slope estimated as cross-spectrum.
the MATLAB chirp function, and the analytic signal was reconstructed using the MATLAB Hilbert transform to estimate the imaginary part of the signal. The STFT was computed using a 129 long Hanning window, with a 75% overlap. The data segments were zero filled to 256 samples. Time derivatives were estimated using a time delay of 1 sample of the original signal. Frequency derivatives were computed with a 1 frequency bin delay. All surfaces were interpolated to ensure that they were computed on the same sampling lattice.

Figure 5 represents the spectrogram with the thresholded remapped STFT surface superimposed. Figure 6 is an expansion of a portion of Fig. 5, showing the remapped points and the granularity of the spectrogram. Figure 7 is the remapped STFT superimposed on the IF computed by differentiating the phase of the reconstructed analytic signal. Of note is the error of the IF estimate at the beginning and the end of the data due to the integration of the Hilbert filter. This error is not present in the remapped STFT. Figure 8 represents an expansion of a portion of Fig. 7. Figures 9 and 10 represent the IS computed as the derivative of the IF and the IS estimated as cross-spectra. Of note is the huge variance in the IF-based estimate computed as the second derivative of the phase of the reconstructed analytic signal. This variance is not evident in the cross-spectrum based estimate.

5. CONCLUSIONS

We have demonstrated an effective method for accurately estimating FM signal components and their derivatives from the STFT phase. The method has been applied to synthesized signals and speech, demonstrating that they are both accurate and effective.

REFERENCES


DOUGLAS NELSON was born in Minneapolis, MN, on 5 November 1945. He received a bachelor’s degree in mathematics from the University of Minnesota in 1967 and a doctorate in mathematics from Stanford University in 1972. After spending three years as an assistant professor at Carnegie-Mellon University, he accepted a position at the National Security Agency at Fort Meade, MD, where he has been from 1975 to the present. At the NSA, Dr. Nelson is a graduate of the Senior Technical Development Program (STDP), a master level member of the Math Tech Track, and the author of more than 100 technical papers. Dr. Nelson’s primary research interest has been developing signal processing algorithms for radar, communications, and speech-related problems. Several of these algorithms have been patented or have patents pending. Nonacademic interests include woodworking, chair making, bagpipe playing, and the restoration of an 18th-century plantation house.